

Multiple Imputation of Missing Data in Educational Production Functions

Amira Elasa

Supplementary materials

Missing value patterns

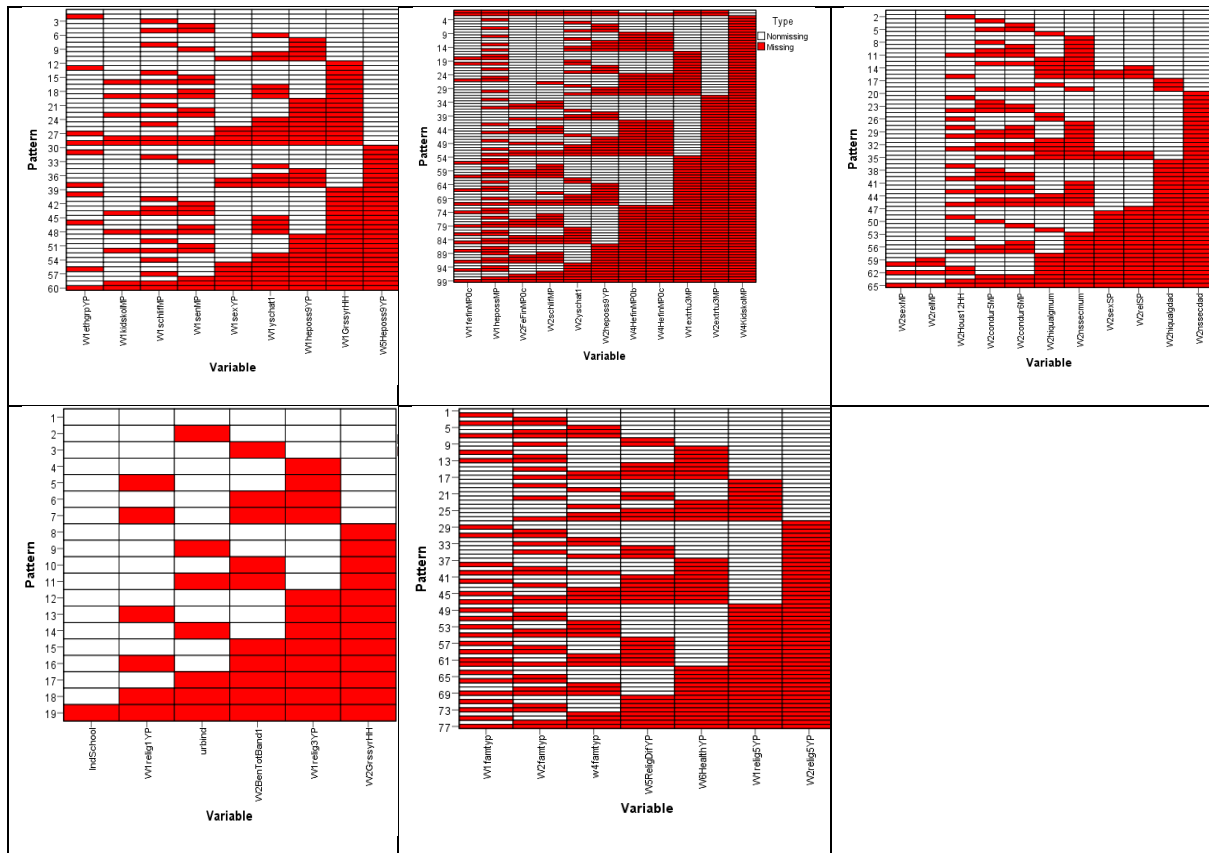


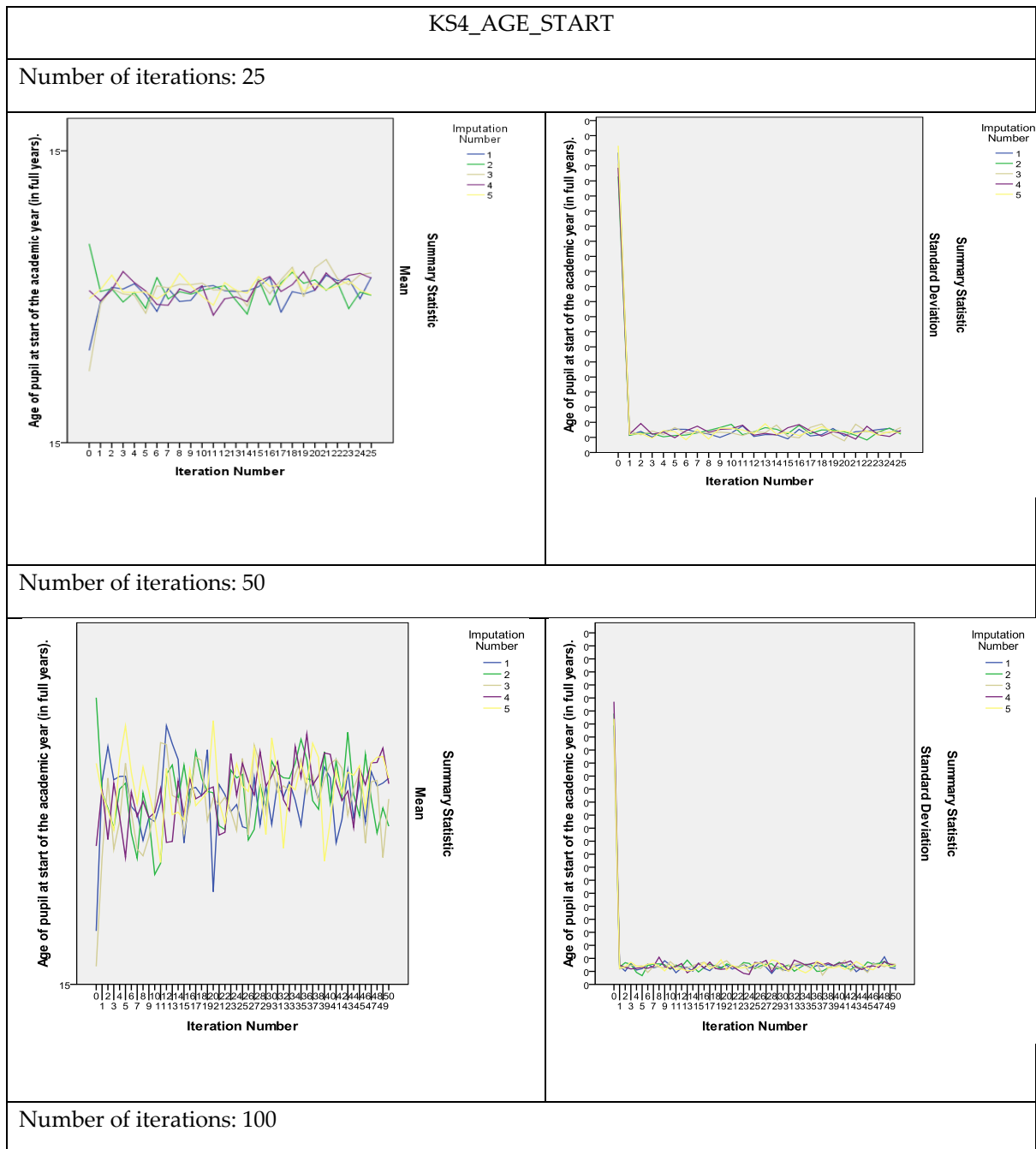
Figure S1. Missing value patterns.

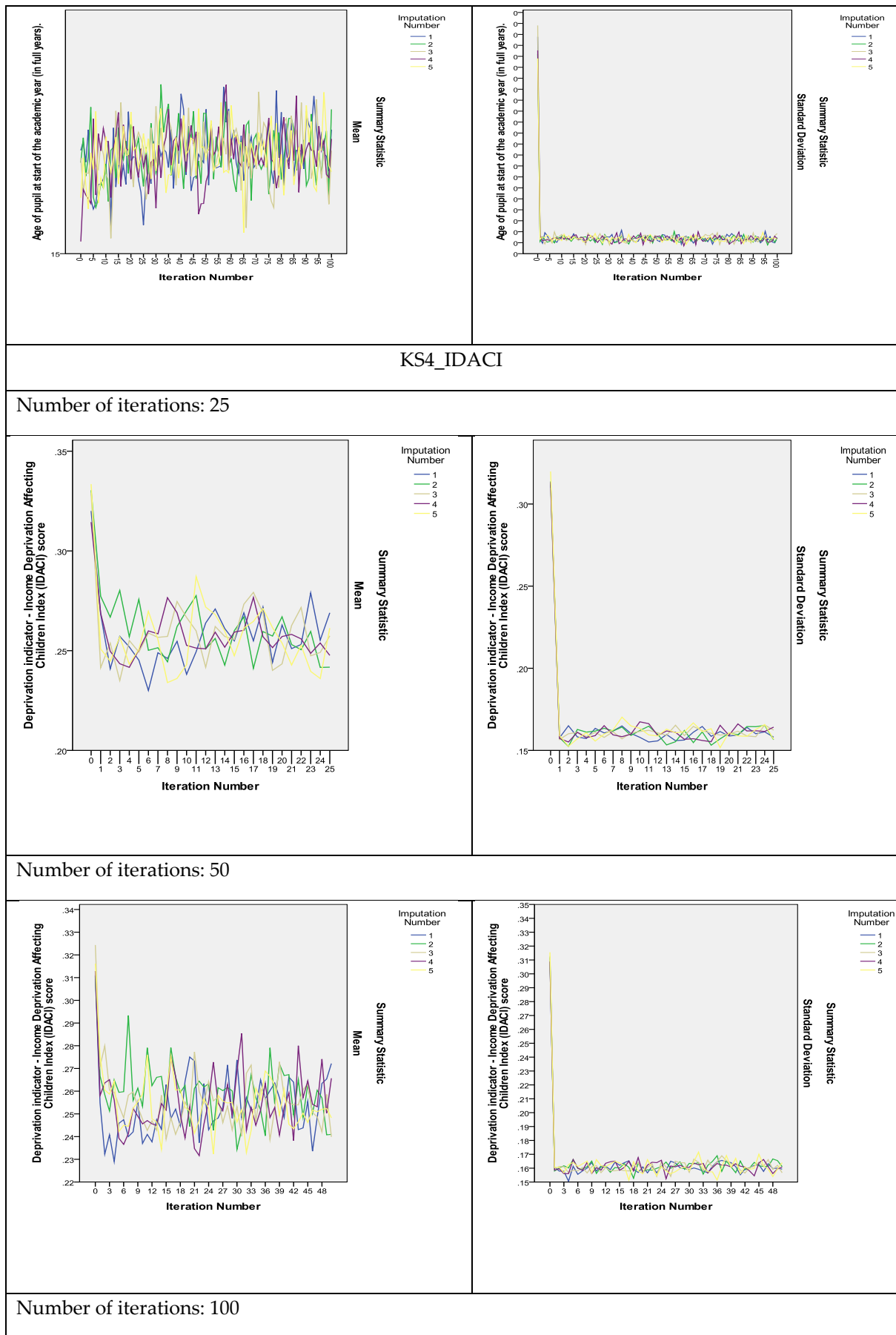
Variable definition

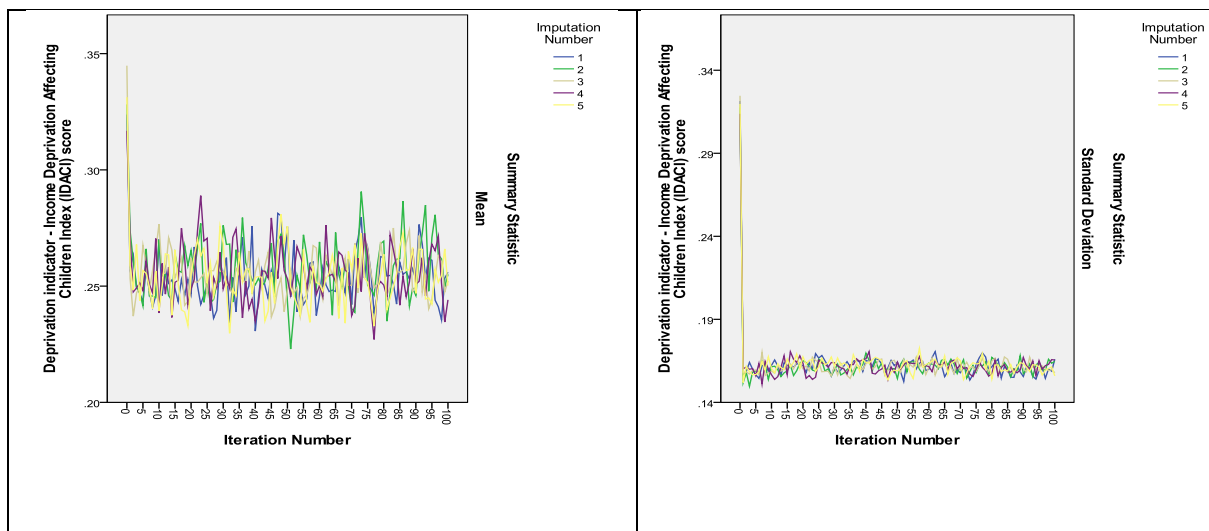
Table S1. Variable definitions

Code	Variable Name
KS4_PASS_AG	KS4 Total number of GCSE qualifications at grades A*-G (GCSE equivalencies).
KS4_PTSTNEWG	Key stage 4 total GCSE/GNVQ new style point score.
W7PayYrMain_Banded1	Average yearly pay for main job (wave 7).
W2BenTotBand1	Banded total amount of benefits (Wave 2).
W1GrssyrHH W2GrssyrHH	Gross annual salary of HH (edited) for wave 1-2.
KS4_CVAP3APS KS4_CVAP2APS	Prior attainment measured by key stage 3 and 2 average point score (using fine grading) for contextual value added.
KS4_AGE_START	Young person's age when started KS4.
KS4_IDACI	Income Deprivation Affecting Children Index.
W1yschat1 W2yschat1	Mean score of the young person's attitude to school for wave 1-2.

Algorithm convergence plots

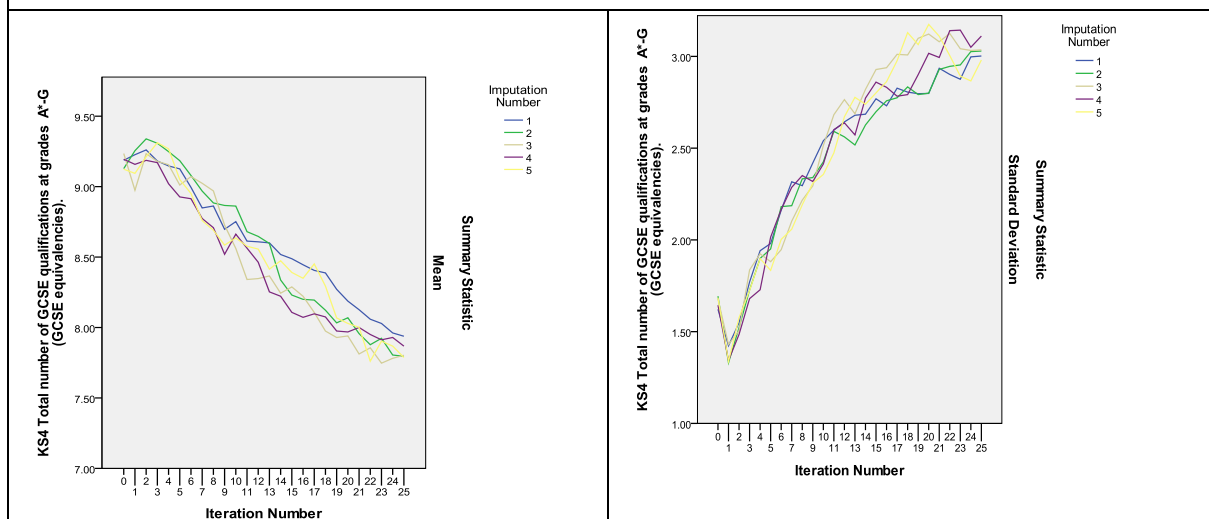




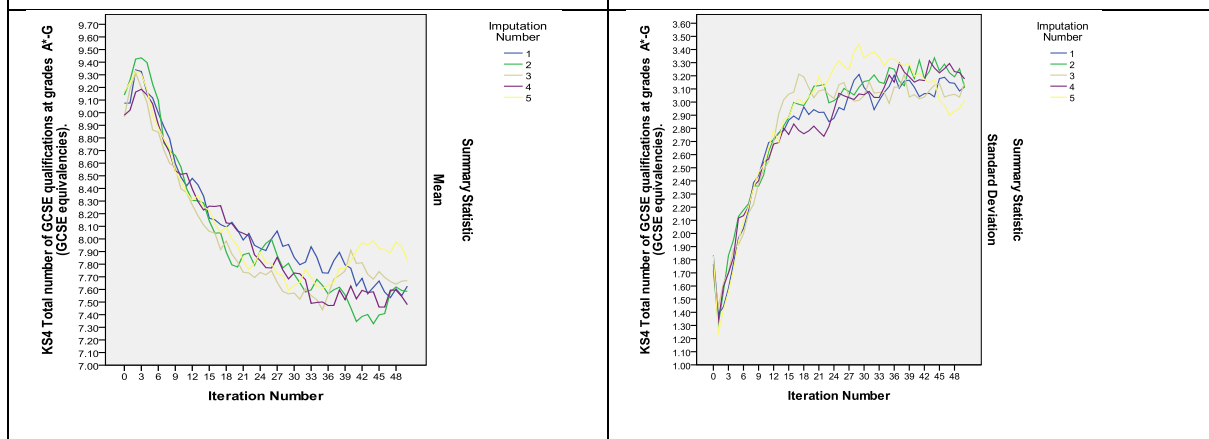


KS4_PASS_AG

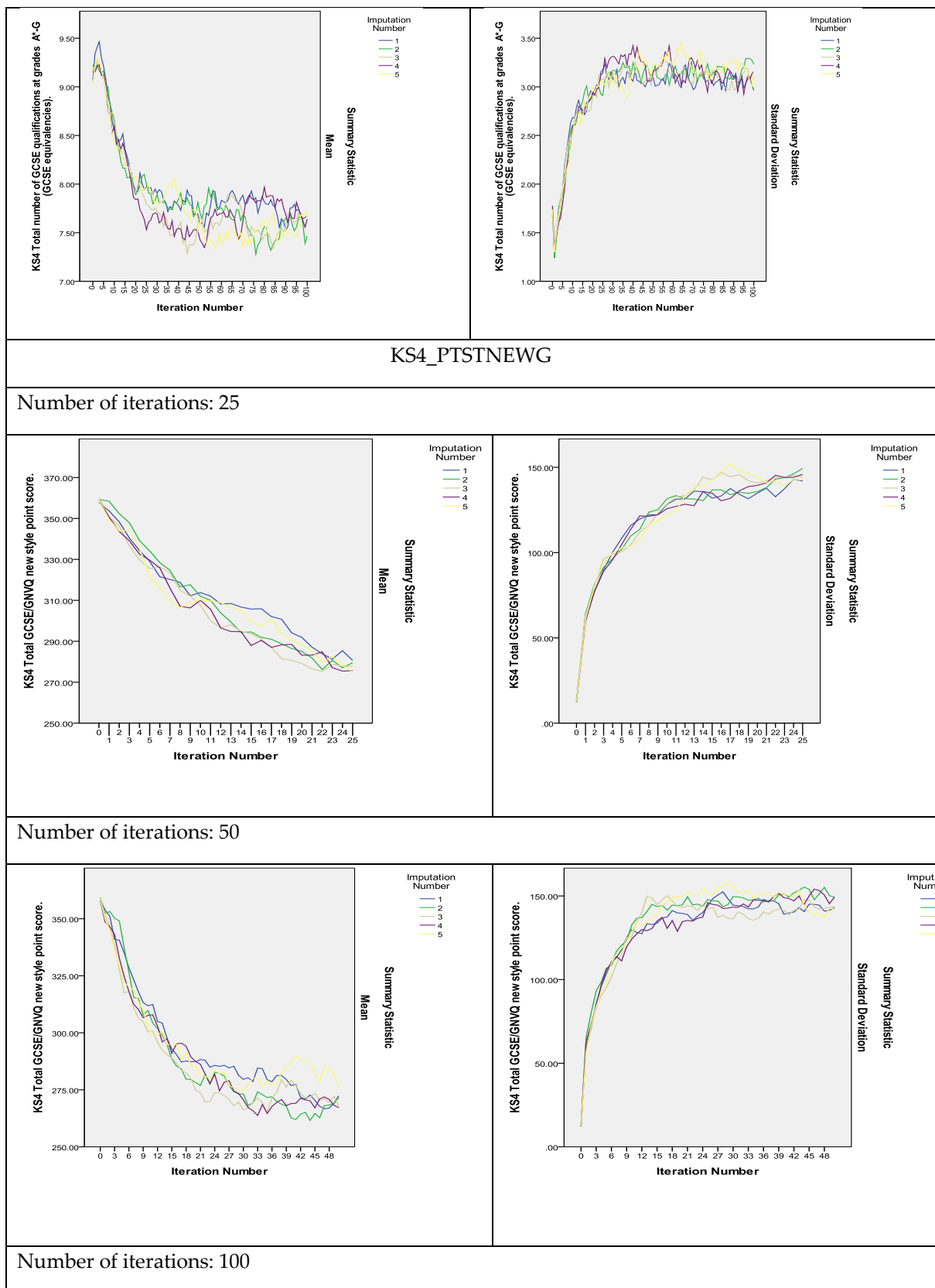
Number of iterations: 25

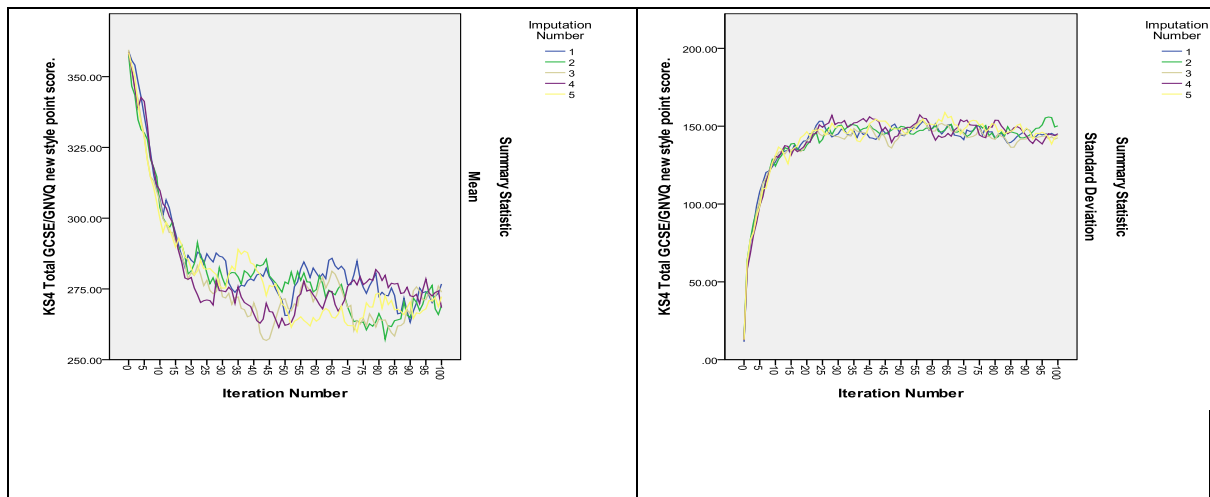


Number of iterations: 50



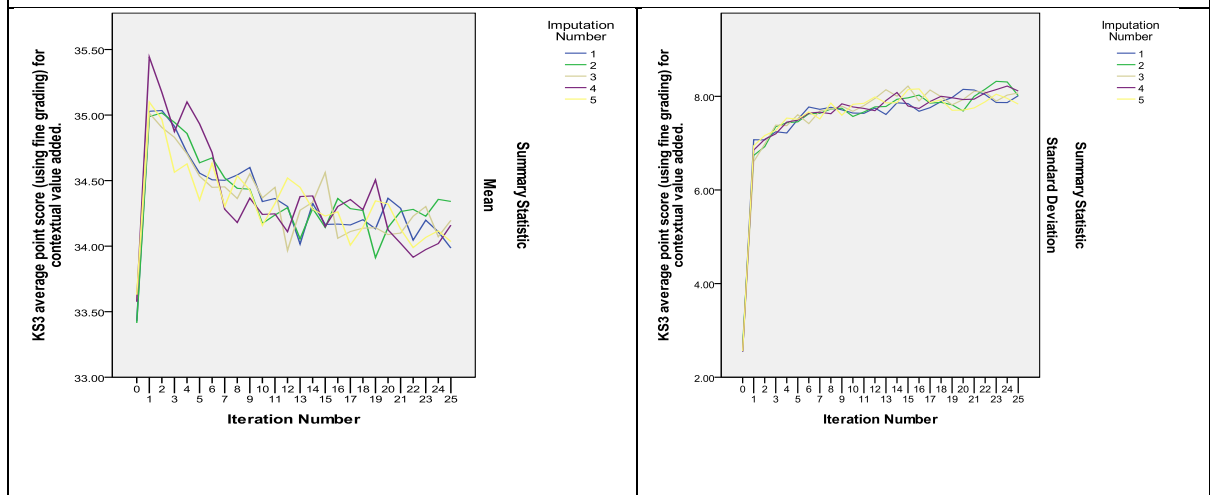
Number of iterations: 100



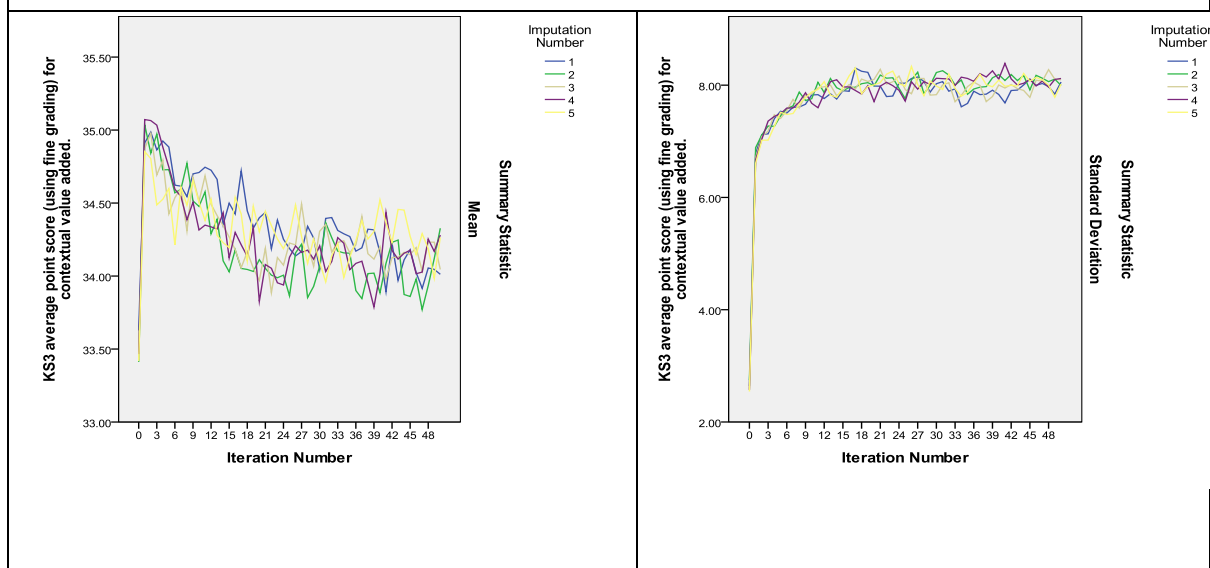


KS4_CVAP3APS

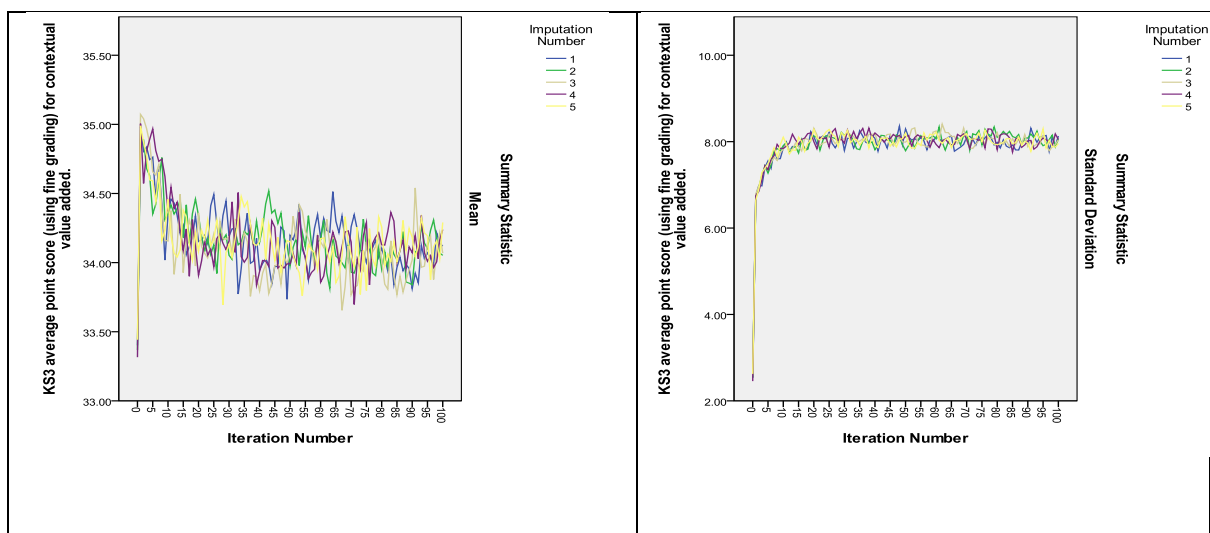
Number of iterations: 25



Number of iterations: 50

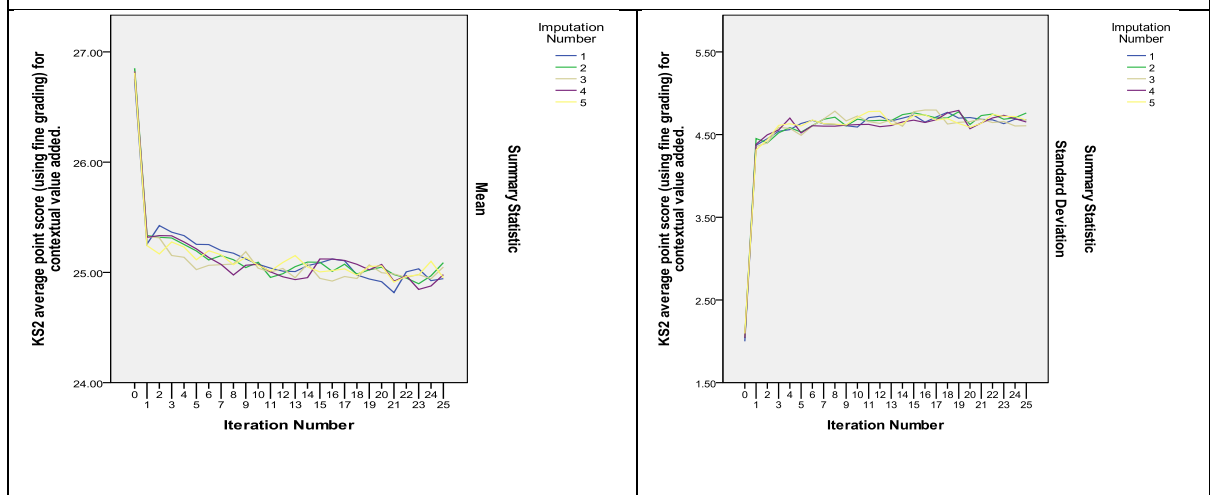


Number of iterations: 100

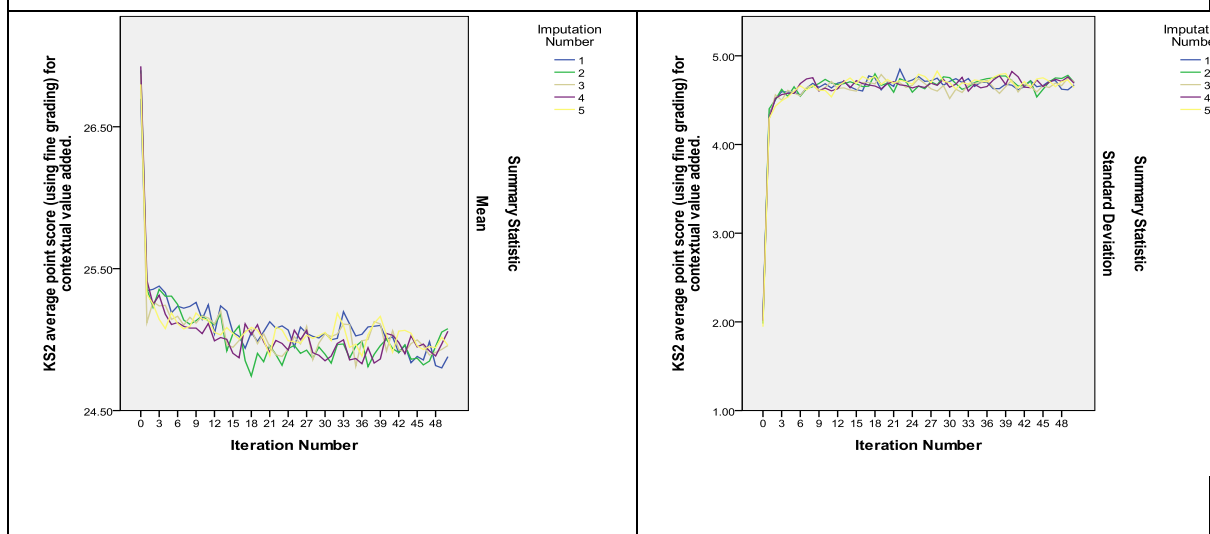


KS4_CVAP2APS

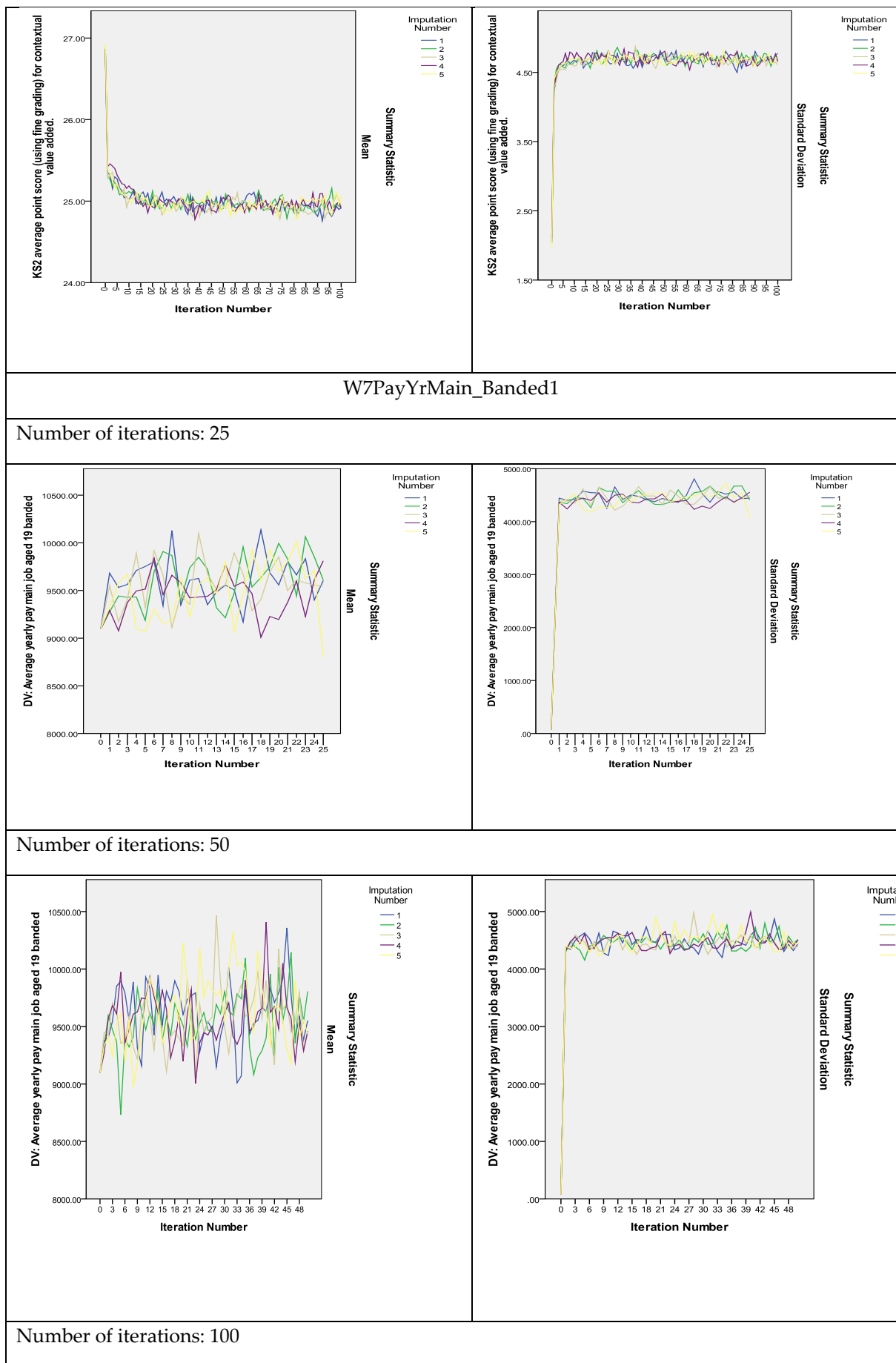
Number of iterations: 25

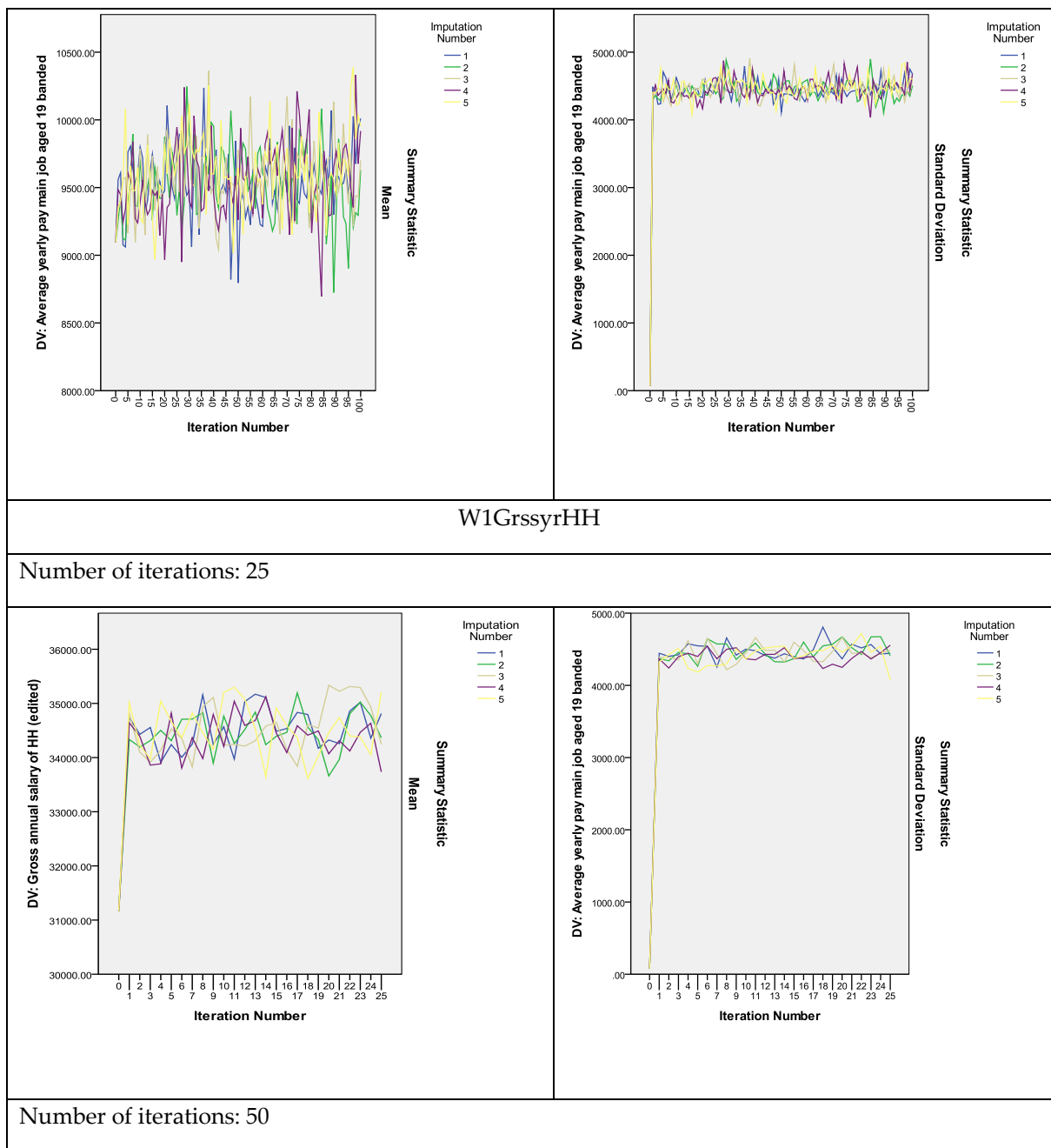


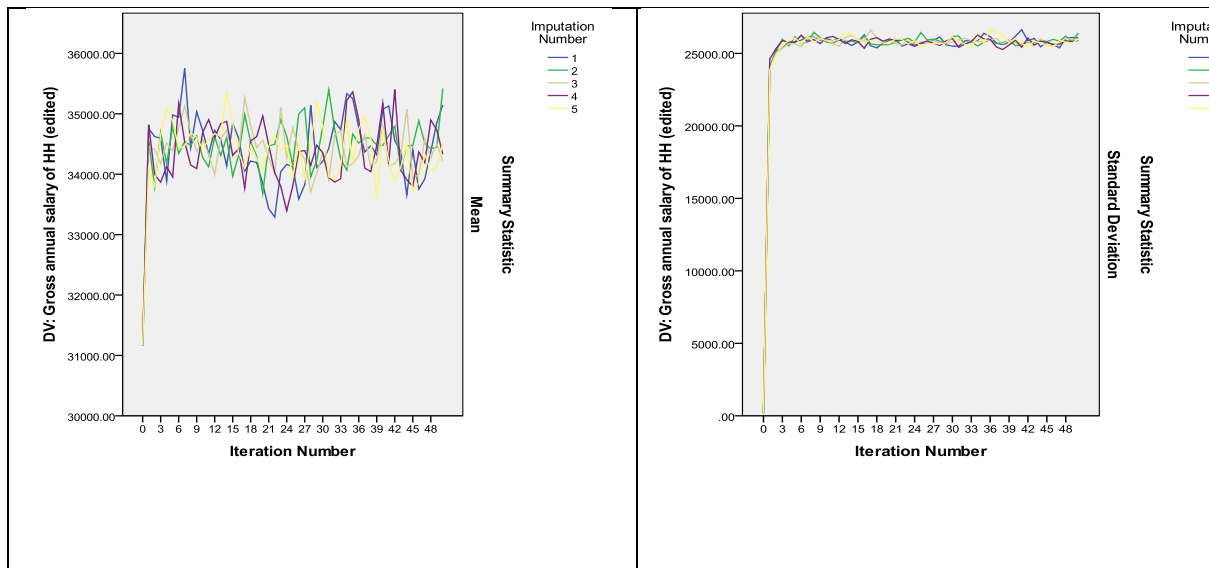
Number of iterations: 50



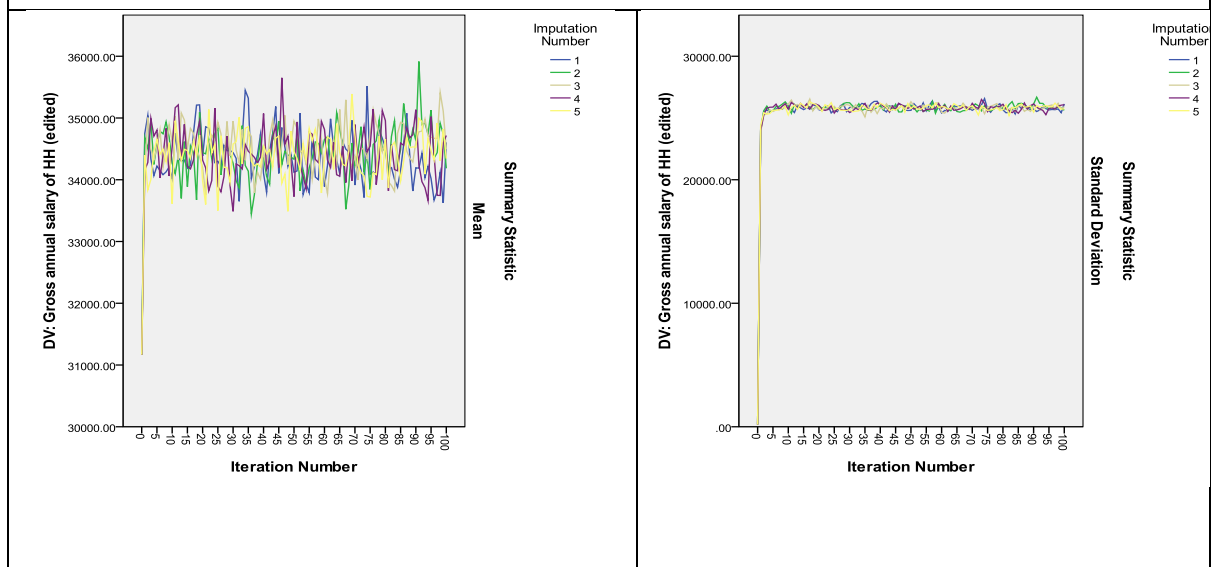
Number of iterations: 100





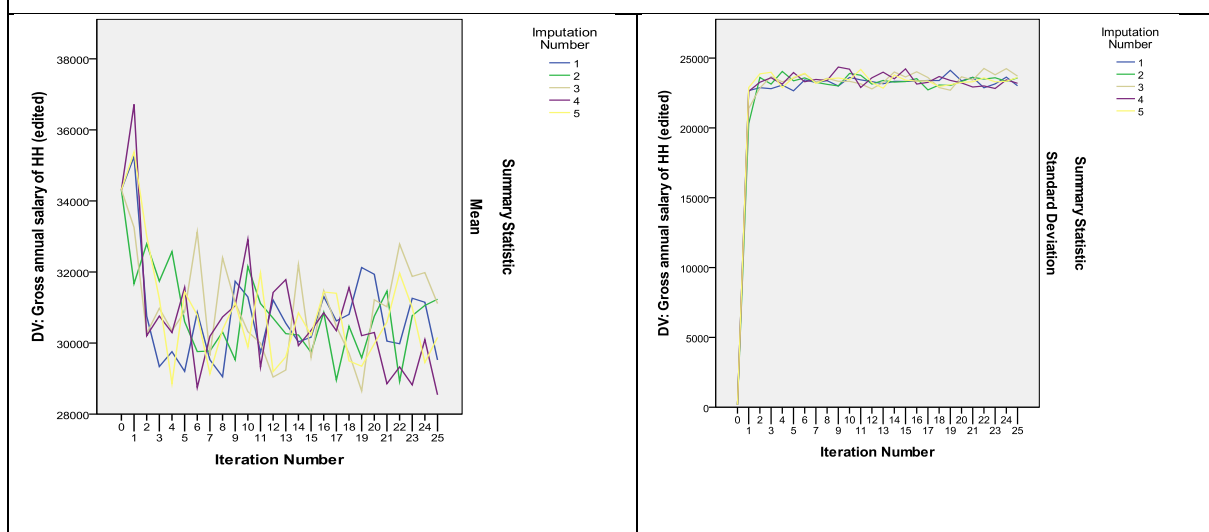


Number of iterations: 100

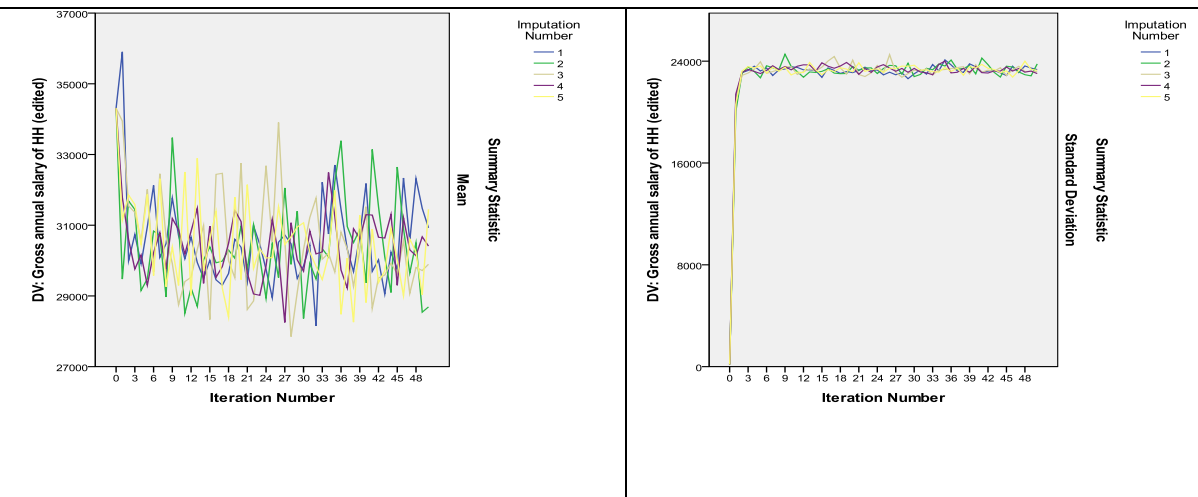


W2GrssyrHH

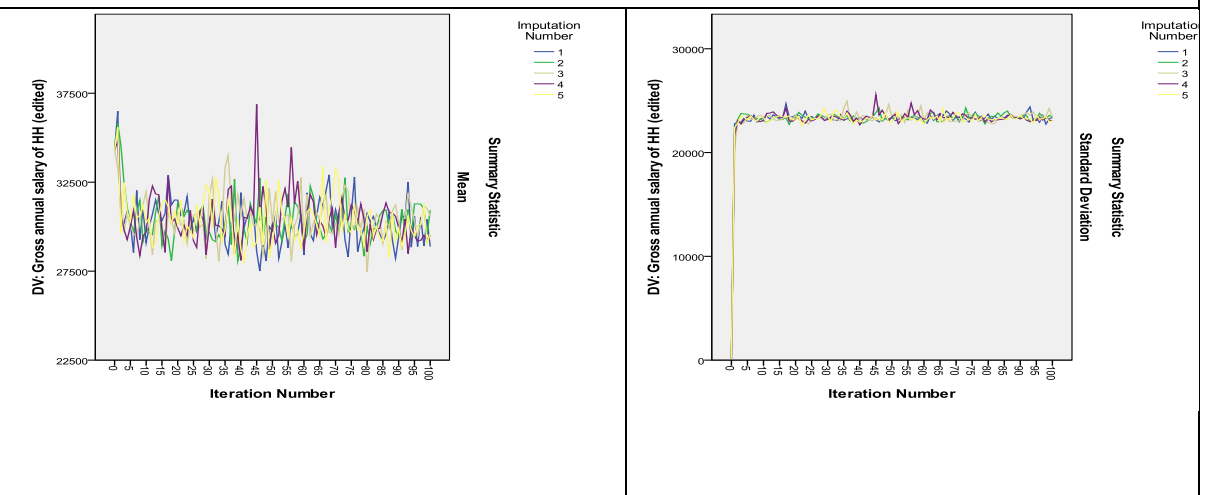
Number of iterations: 25



Number of iterations: 50

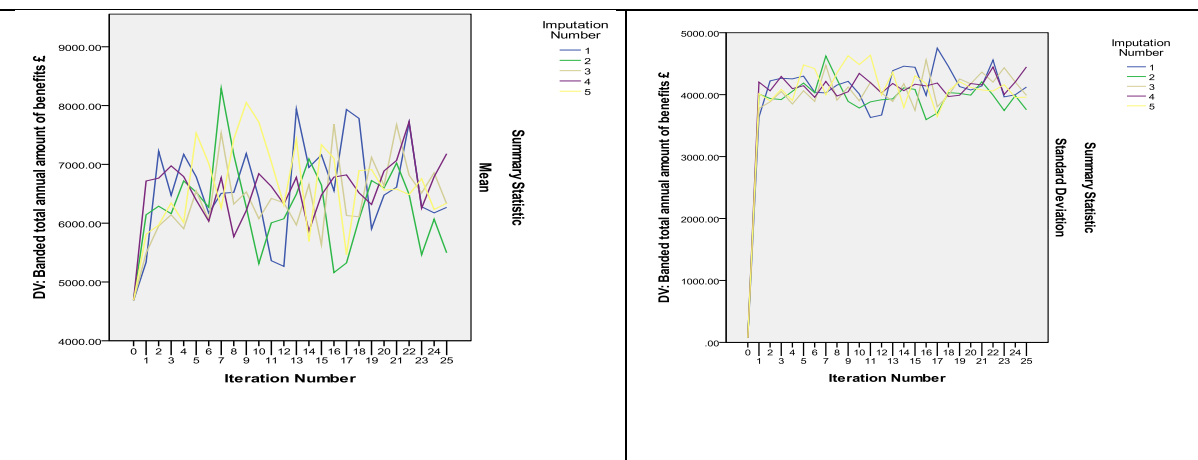


Number of iterations: 100

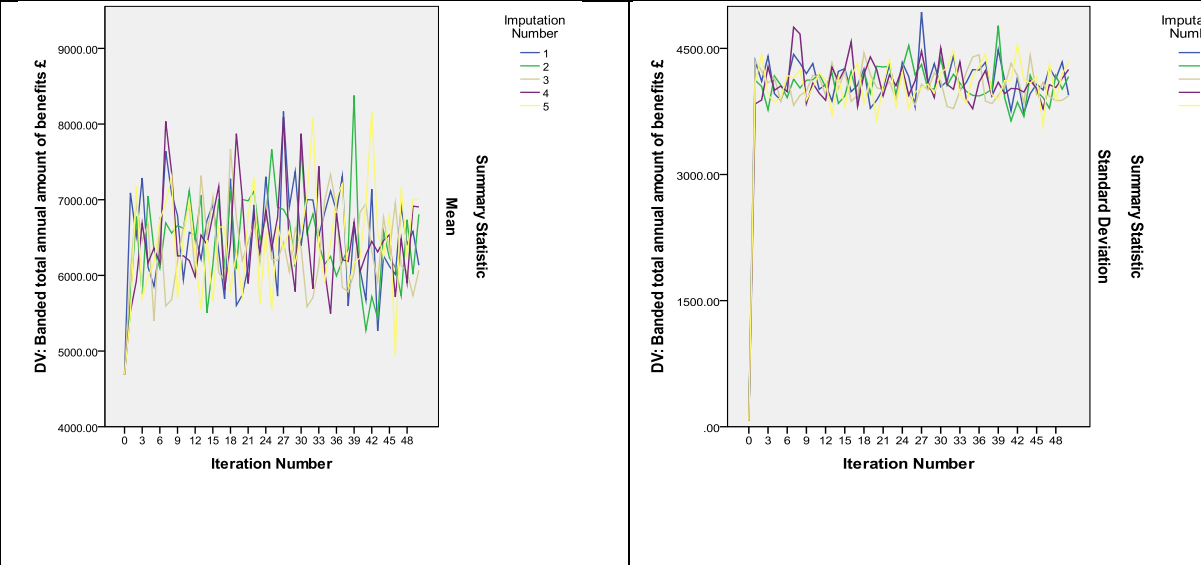


W2BenTotBand1

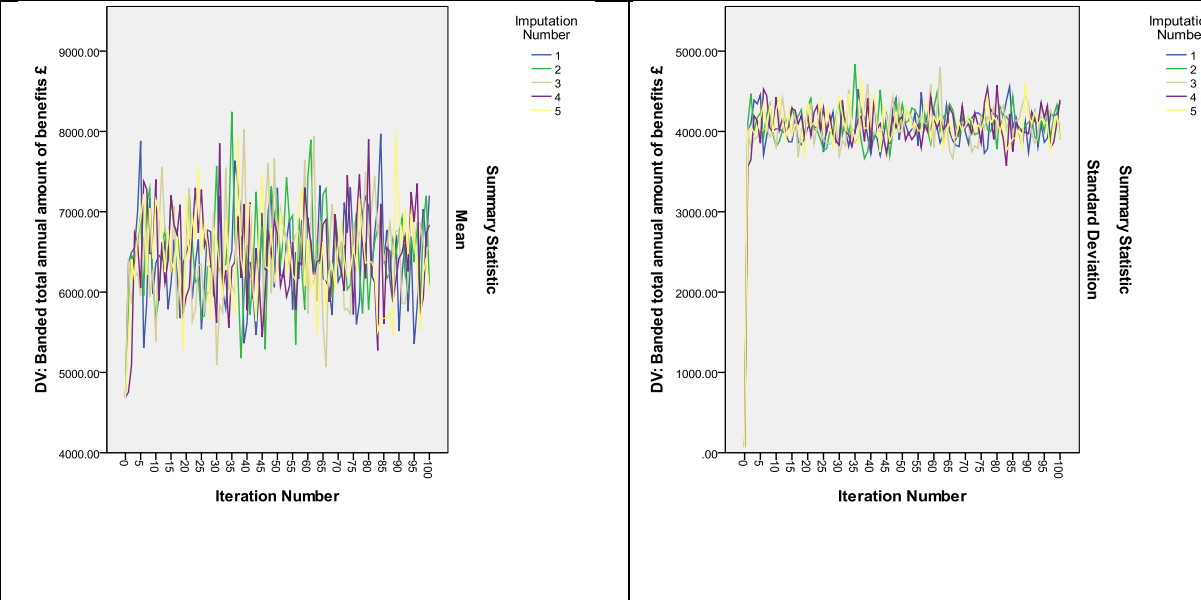
Number of iterations: 25



Number of iterations: 50

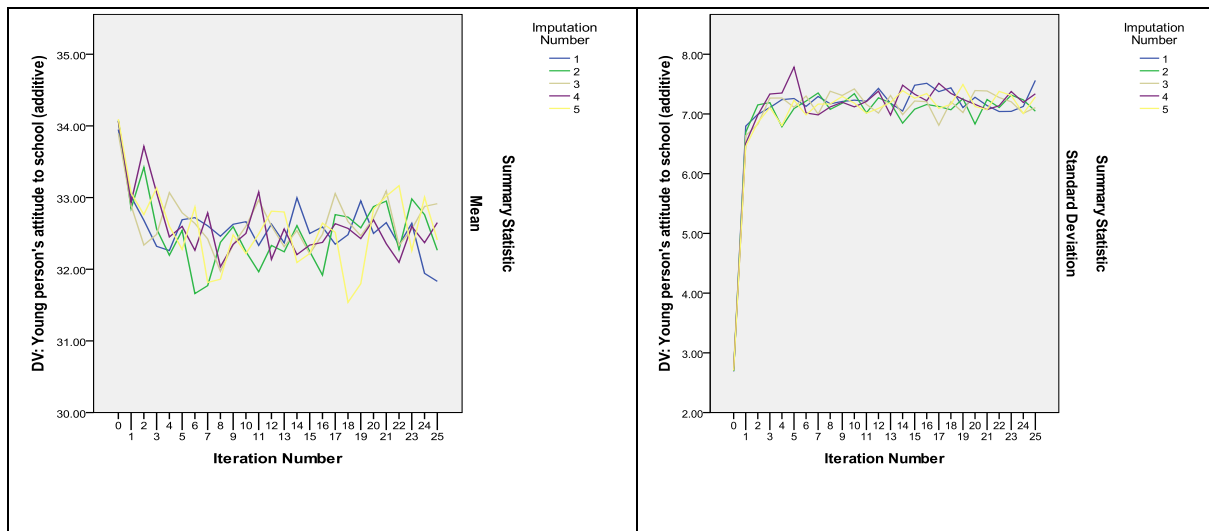


Number of iterations: 100

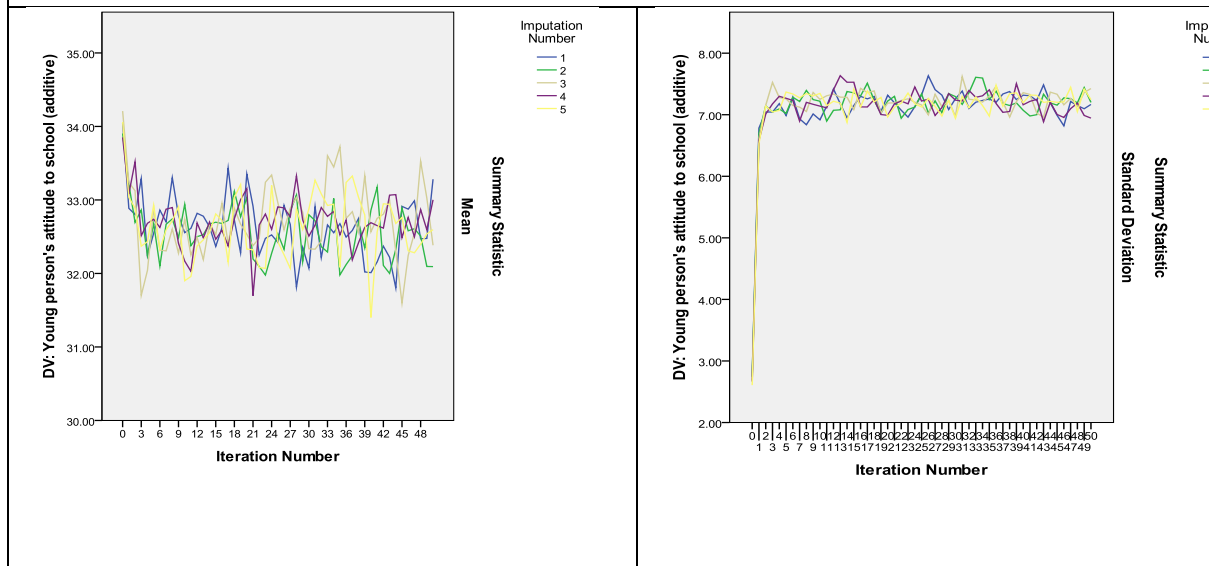


W1yschat1

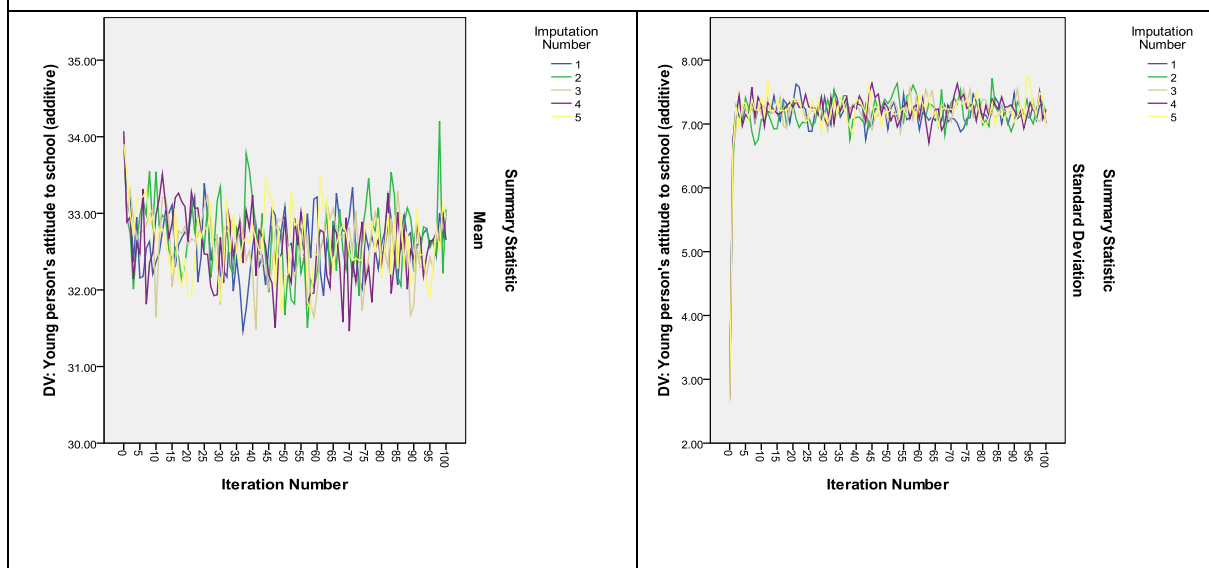
Number of iterations: 25



Number of iterations: 50

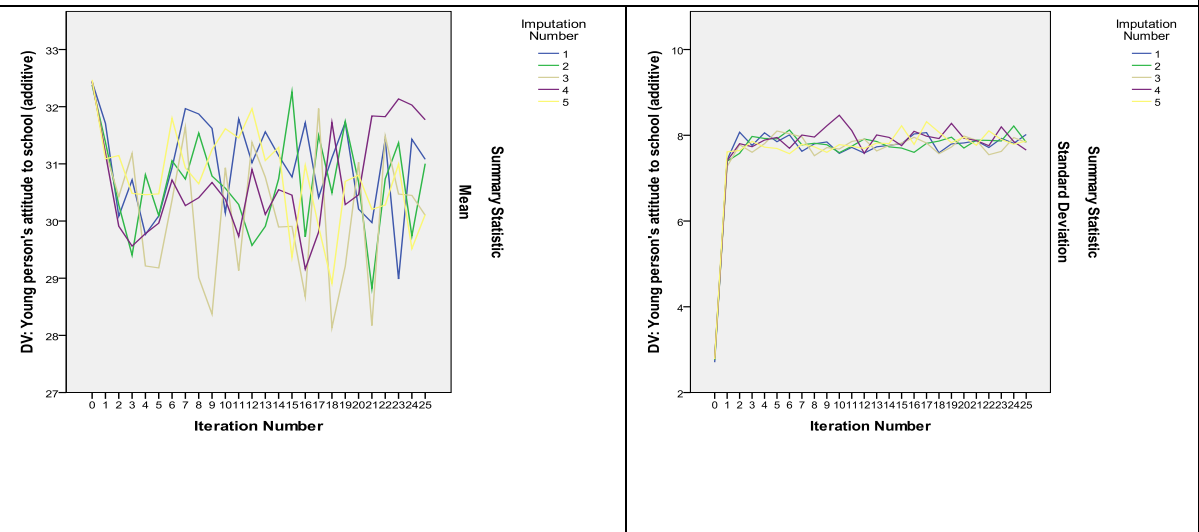


Number of iterations: 100

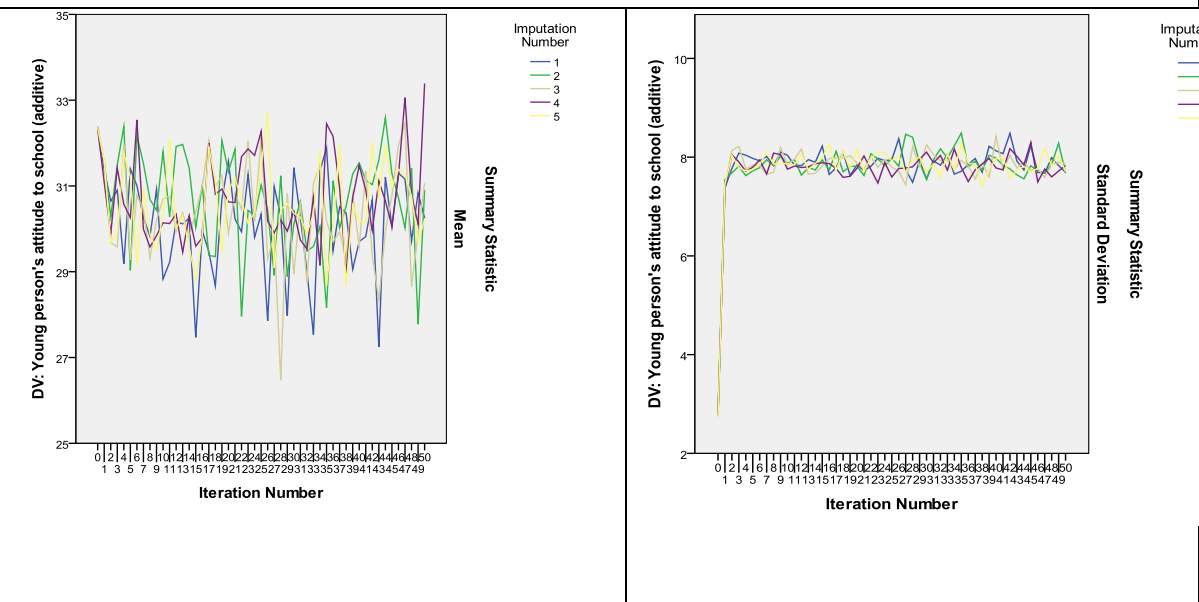


W2yschat1

Number of iterations: 25



Number of iterations: 50



Number of iterations: 100

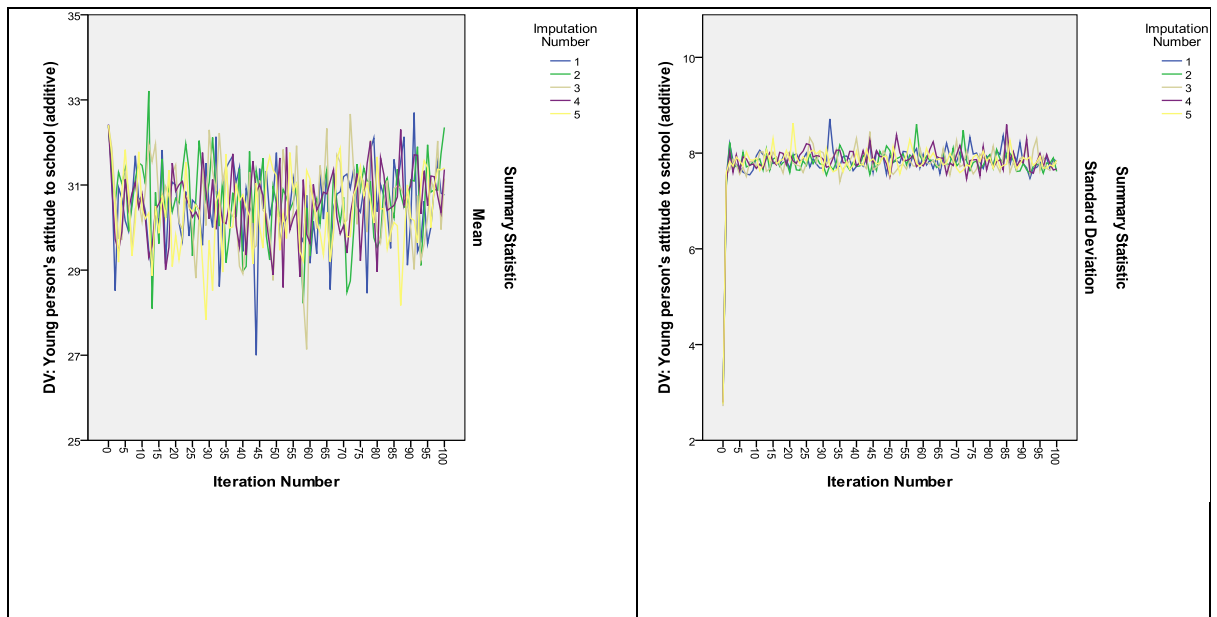
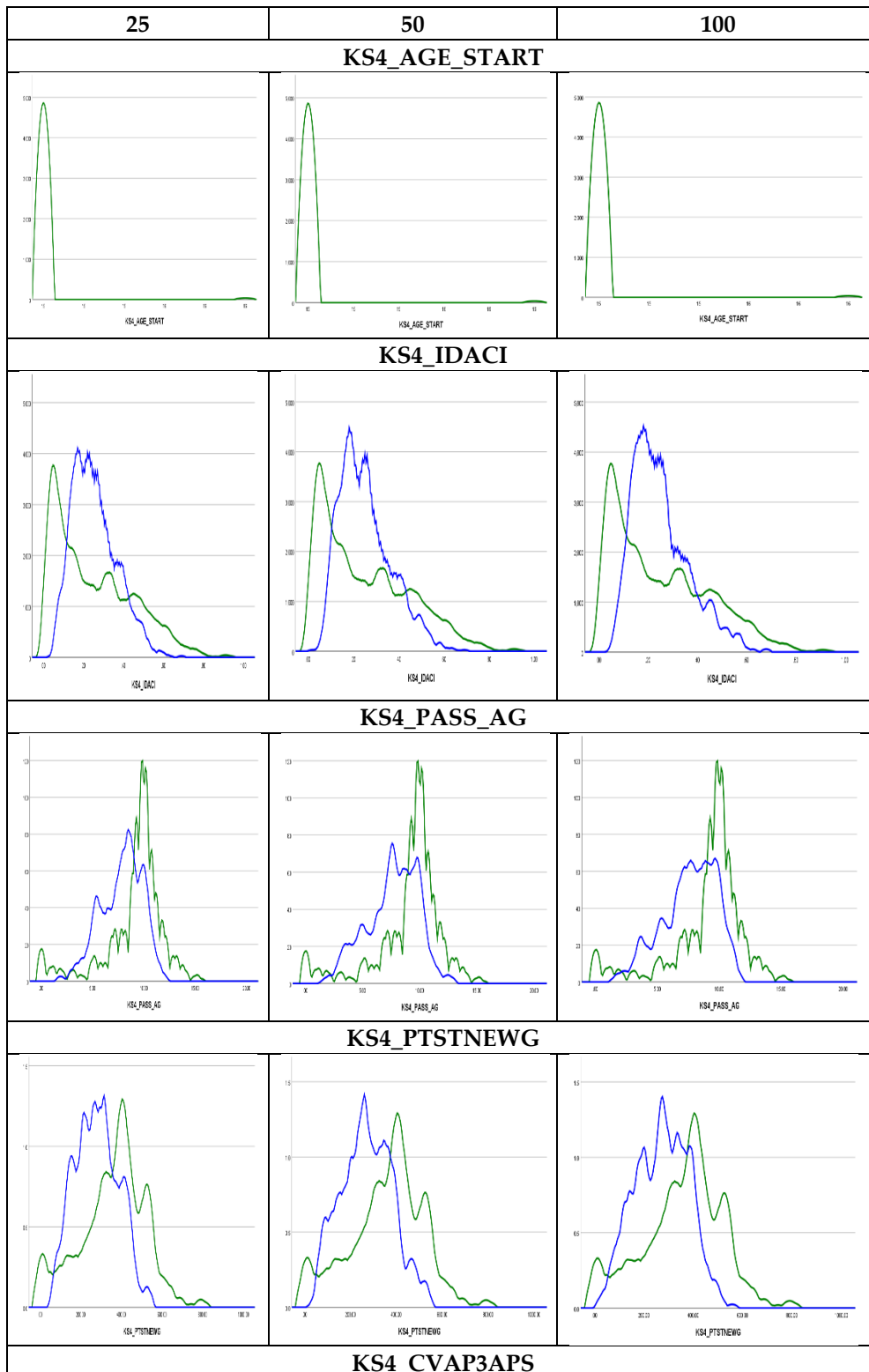
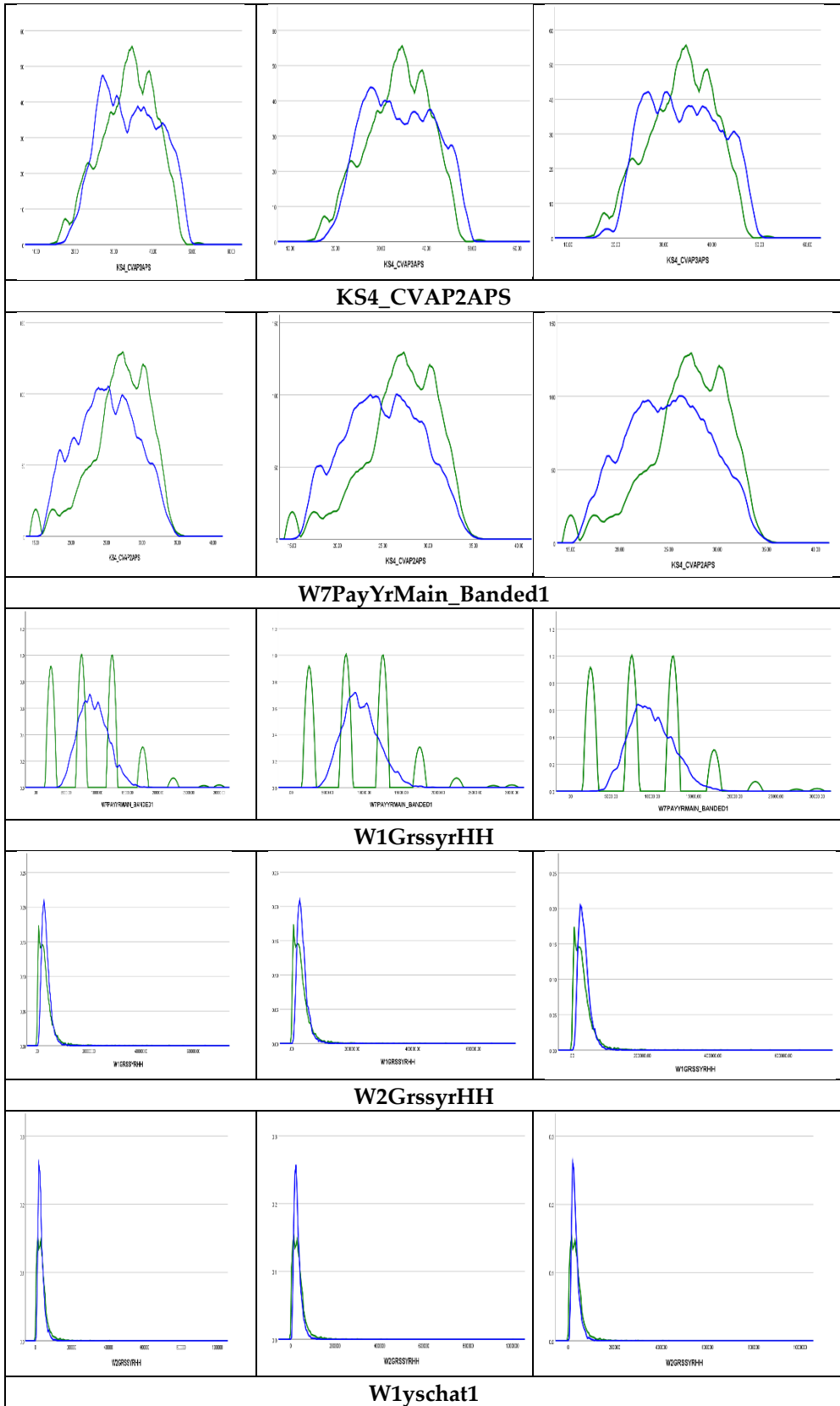


Figure S2. Convergence of the MI algorithm

Kernel density plots





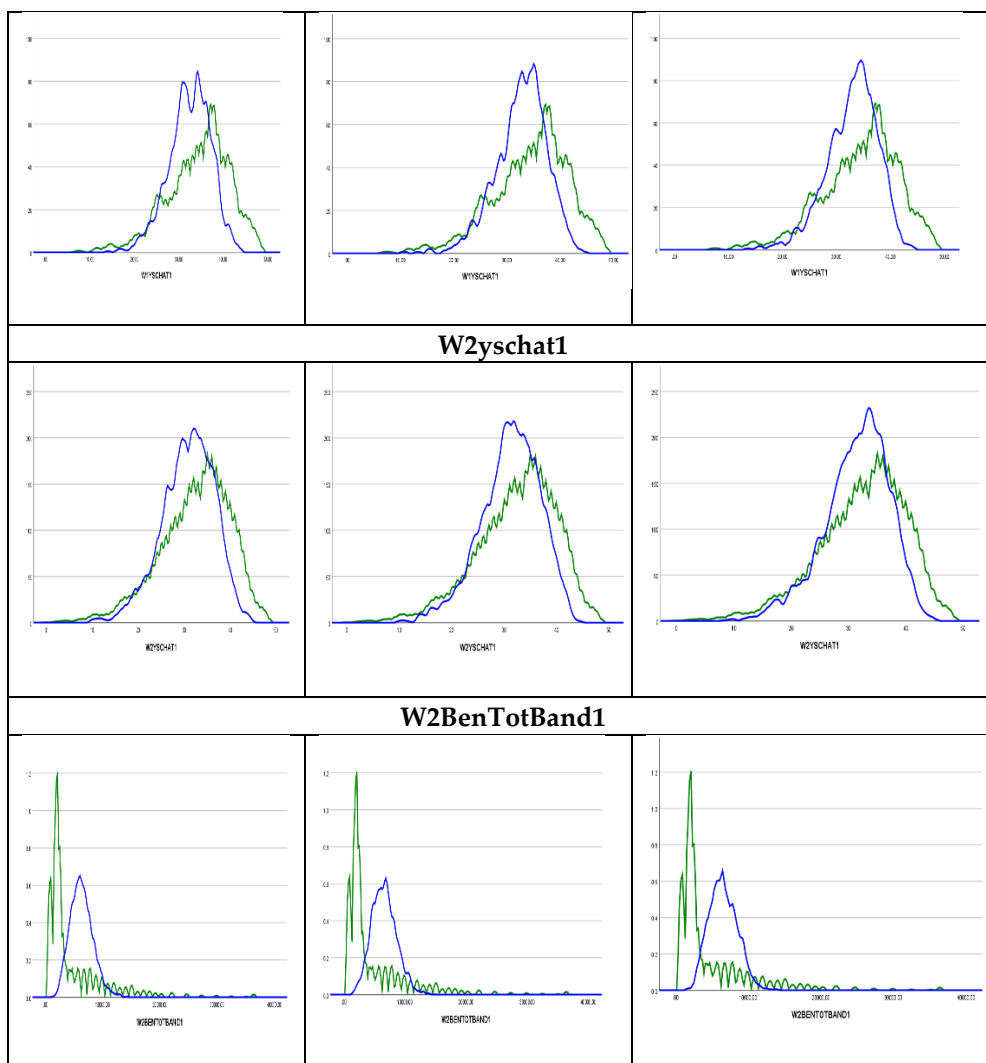


Figure S3. Kernel density plots. Observed values are presented by the green curve and the imputed values are presented by blue curve.

Kolmogorov-Smirnov (KS) tests

Table S2. Kolmogorov-Smirnov (KS) test p-values at 10%, 5% and 1% significance levels.

Variable/iterations	25		50		100	
	Observed	Imputed	Observed	Imputed	Observed	Imputed
KS4_AGE_START						
10%	0.000	0.200	0.000	0.165	0.000	0.162
5%	0.000	0.200	0.000	0.165	0.000	0.162
1%	0.000	0.200	0.000	0.165	0.000	0.162
KS4_IDACI						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
KS4_PASS_AG						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
KS4_PTSTNEWG						
10%	0.000	0.098	0.000	0.098	0.000	0.0846
5%	0.000	0.098	0.000	0.098	0.000	0.0846
1%	0.000	0.098	0.000	0.098	0.000	0.0846
KS4_CVAP3APS						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
KS4_CVAP2APS						
10%	0.000	0.000	0.000	0.000	0.000	0.0006
5%	0.000	0.000	0.000	0.000	0.000	0.0006
1%	0.000	0.000	0.000	0.000	0.000	0.0006
W7PayYrMain_Banded1						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
W1GrssyrHH						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
W2GrssyrHH						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000
W1yschat1						
10%	0.000	0.021	0.000	0.039	0.000	0.0178
5%	0.000	0.021	0.000	0.039	0.000	0.0178
1%	0.000	0.021	0.000	0.039	0.000	0.0178
W2yschat1						
10%	0.000	0.002	0.000	0.007	0.000	0.0008
5%	0.000	0.002	0.000	0.005	0.000	0.0008

1%	0.000	0.002	0.000	0.005	0.000	0.0008
W2BenTotBand1						
10%	0.000	0.000	0.000	0.000	0.000	0.000
5%	0.000	0.000	0.000	0.000	0.000	0.000
1%	0.000	0.000	0.000	0.000	0.000	0.000

Absolute difference of means between observed and imputed data

Table S3. Absolute difference of means between observed and imputed data

Variable Name/iterations	Absolute difference in means	2 standard deviations (25)	2 standard deviations (50)	2 standard deviations (100)
KS4_AGE_START	0.000	0.992	0.992	0.992
KS4_IDACI	0.008	1.676	1.678	1.682
KS4_PASS_AG	1.278	6.915	6.961	6.965
KS4_PTSTNEWG	81.211	49.311	49.365	49.346
KS4_CVAP3APS	0.641	10.868	10.886	10.882
KS4_CVAP2APS	1.849	8.386	8.390	8.392
W7PayYrMain_Banded1	381.176	281.911	282.786	284.969
W1GrssyrHH	3307.491	675.833	676.990	676.160
W2GrssyrHH	4197.773	656.329	655.701	655.478
W2BenTotBand1	1632.950	273.261	274.299	274.307
W1yschat1	1.632	10.796	10.770	10.728
W2yschat1	1.582	11.116	11.085	11.082

Ratios of variances

Table S4. Ratios of variances of observed and imputed data

Variable Name/iterations	25	50	100
KS4_AGE_START	1.050	1.050	1.050
KS4_IDACI	1.201	1.195	1.183
KS4_PASS_AG	0.971	0.946	0.944
KS4_PTSTNEWG	1.099	1.094	1.096
KS4_CVAP3APS	0.843	0.838	0.839
KS4_CVAP2APS	0.881	0.879	0.878
W7PayYrMain_Banded1	1.260	1.243	1.201
W1GrssyrHH	1.209	1.200	1.207
W2GrssyrHH	1.299	1.305	1.307
W2BenTotBand1	1.302	1.280	1.279
W1yschat1	1.005	1.015	1.031
W2yschat1	0.970	0.980	0.981

P-values of t-tests of means differences between observed and pooled imputed values

Table S5. P-values of t-tests of means differences between observed and pooled imputed values +

Variable/iterations/significance level	25	50	100
KS4_AGE_START			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
KS4_IDACI			
10%	0.000	0.000	0.003
5%	0.000	0.000	0.003
1%	0.000	0.000	0.003
KS4_PASS_AG			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
KS4_PTSTNEWG			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
KS4_CVAP3APS			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
KS4_CVAP2APS			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W7PayYrMain_Banded1			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W1GrssyrHH			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W2GrssyrHH			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W1yschat1			

10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W2yschat1			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000
W2BenTotBand1			
10%	0.000	0.000	0.000
5%	0.000	0.000	0.000
1%	0.000	0.000	0.000

+The pooled p-values are calculated using the mean of the p-values for the five imputed values.

F-test and t-tests results

Table S6. F-test and t-tests p-values*

Variable	t test p-values			F test p-values		
Iterations/imputation	25	50	100	25	50	100
KS4_AGE_START						
1	0.537	0.537	0.537	0.069	0.069	0.069
2	0.537	0.537	0.537	0.069	0.069	0.069
3	0.537	0.537	0.537	0.069	0.069	0.069
4	0.537	0.537	0.537	0.069	0.069	0.069
5	0.537	0.537	0.537	0.069	0.069	0.069
KS4_IDACI						
1	0.000	0.000	0.106	0.000	0.000	0.000
2	0.271	0.22	0.152	0.000	0.000	0.000
3	0.055	0.137	0.344	0.000	0.000	0.000
4	0.958	0.000	0.521	0.000	0.000	0.000
5	0.006	0.869	0.591	0.000	0.000	0.000
KS4_PASS_AG						
1	0.000	0.000	0.000	0.003	0.000	0.002
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.001	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.005	0.001	0.000
KS4_PTSTNEWG						
1	0.000	0.000	0.000	0.019	0.168	0.11
2	0.000	0.000	0.000	0.965	0.283	0.427
3	0.000	0.000	0.000	0.048	0.073	0.172
4	0.000	0.000	0.000	0.096	0.649	0.23
5	0.000	0.000	0.000	0.062	0.067	0.089
KS4_CVAP3APS						
1	0.073	0.060	0.024	0.000	0.000	0.000
2	0.002	0.002	0.043	0.000	0.000	0.000
3	0.011	0.096	0.037	0.000	0.000	0.000
4	0.016	0.004	0.007	0.000	0.000	0.000
5	0.044	0.005	0.003	0.000	0.000	0.000
KS4_CVAP2APS						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.000	0.000	0.000
W7PayYrMain_Banded1						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000

3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.002	0.000	0.000	0.000	0.000	0.000
W1GrssyrHH						
1	0.000	0.000	0.000	0.143	0.094	0.010
2	0.000	0.000	0.000	0.165	0.011	0.367
3	0.000	0.000	0.000	0.194	0.183	0.168
4	0.000	0.000	0.000	0.693	0.303	0.095
5	0.000	0.000	0.000	0.032	0.184	0.356
W2GrssyrHH						
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.002	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.000	0.000	0.001	0.000	0.000
W1yschat1						
1	0.000	0.002	0.000	0.027	0.787	0.758
2	0.000	0.000	0.000	0.897	0.628	0.475
3	0.000	0.000	0.000	0.663	0.104	0.658
4	0.000	0.000	0.000	0.226	0.486	0.766
5	0.000	0.000	0.000	0.388	0.369	0.203
W2yschat1						
1	0.000	0.000	0.000	0.000	0.321	0.005
2	0.000	0.000	0.806	0.002	0.111	0.138
3	0.000	0.000	0.000	0.01	0.000	0.020
4	0.000	0.000	0.000	0.064	0.000	0.113
5	0.000	0.000	0.000	0.003	0.100	0.005
W2BenTotBand1						
1	0.000	0.000	0.000	0.000	0.000	0.005
2	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.000	0.000	0.027	0.000	0.002
5	0.000	0.000	0.000	0.000	0.001	0.000

+ The p-values are reported for tests at 5% significance levels. The tests results for 10% and 1% significance levels are identical.